Coexistence of Strategies and Culturally-specific Common Knowledge: an Evolutionary Analysis

*Angelo Antoci**

*Pier Luigi Sacco***

*Luca Zarri****

**Department of Economics and Business, University of Sassari*

***Department of Economics, University of Bologna*

****School of Economic and Social Studies, University of East Anglia*

Synopsis

We analyze social dynamics in a continuous population where randomly matched individuals have to choose between two pure strategies only ('cooperate' (C) and 'not cooperate' (NC)). Individual payoffs associated to the possible outcomes of each interaction may differ across groups, depending on the specific social and cultural context to which each agent belongs. In particular, it is assumed that three sub-populations are initially present, 'framing' the game according to the Prisoner's Dilemma (PD), Assurance Game (AG) and Other Regarding (OR) payoff configurations respectively. In this context, we examine both the adoption process of strategies C and NC within each sub-population and the diffusion process of 'types' (PD, AG and OR) within the overall community. On the basis of an evolutionary game-theoretic approach, the paper focuses on the problem of *coexistence* of PD, AG and OR groups as well as of "nice" (C) and "mean" (NC) strategies. In particular, we show that coexistence between C and NC is possible in the heterogeneous community under examination, even if it is ruled out in homogeneous communities where only one of the three types is present.

Keywords: Prisoner's Dilemma, Social dynamics, Evolutionary game theory.

JEL Classification: C79; Z13

Introduction

Though the Prisoner's Dilemma has been extensively studied under a wide variety of conditions and perspectives (see e.g. Kandori 1982, Rubinstein 1986, Binmore & Samuelson 1992, Ellison 1994), *coexistence* of strategies has rarely been obtained as a theoretical result. Eshel et al. (1999) have considered a large population with a local interaction structure, where unrelated individuals often meet with their neighbors and are allowed to occasionally change their own strategy, by imitating the most successful agents belonging to the interaction neighborhood. In this framework, they define as 'unbeatable' a strategy which turns out to be *robust* against the invasion of a finite group of identical mutants and find that, whenever agents play either Prisoner's Dilemma or Chicken game, cooperation is the *unique* unbeatable strategy insofar as the learning neighborhood is far larger than the interaction neighborhood.

Under very different conditions, Karandikar et al. (1998) obtain a somewhat similar conclusion in a model where two agents play the Prisoner's Dilemma over time and follow an aspiration-based adjustment rule: such a process leads to the eventual emergence of the mutual cooperation outcome (i.e. even in this framework, coexistence is ruled out). However, Palomino & Vega-Redondo (1999) correctly point out that such a result crucially depends on the presence of inter-agent 'feedback effects' due to the small number of players involved in the game. On the contrary, in their paper they set up an aspiration-based dynamic model of bounded rationality where a continuum of agents are randomly matched and play the Prisoner's Dilemma: under certain conditions, their analysis brings about an interesting coexistence result, as long-run *partial* cooperation (never exceeding half of the population) emerges as the unique limit outcome of social adjustment paths. Hirshleifer & Martinez Coll (1991) analyse the dynamics related to the adoption process of four pure strategies ('cooperate', 'tit for tat', 'defect' and 'bully') within a large population of utility-maximising agents. The subjects have to choose which strategy to adopt in a series of random pairwise matchings with other individuals belonging to the same population. The four strategies are played both with payoff configurations of Prisoner's Dilemma (PD) type and of Chicken game (CG) type.

However, it is assumed that the two games are played *separately*: they first consider the adoption process in a population where all the agents believe they are playing a PD game (whose payoff levels are known to all) and are all rationally maximising their own payoff; subsequently, the same process takes place within a population where agents play a CG and they believe this information is *common knowledge*.

In such a context, the adoption process of the above behavioral options leads to the coexistence of 'nice' strategies (such as 'cooperate' and 'tit for tat') and 'mean' strategies (such as 'defect' and 'bully'). As a consequence, their predictions are consistent with the following, well-known experimental[1] and empirical result: despite (normally relevant) cultural and economic differences, in many *large* social environments, a *mixture* of 'nice' and 'mean' behaviors is often observed, as almost everywhere some people are honest and, say, tend to return valuable lost items, to tip in restaurants and to queue in the markets, whereas some other people belonging to the same population do not[2].

The same holds even for more proactive and morally demanding behaviors such as volunteering, contributing to charities, donating blood without monetary reward, voting and saving unknown people at the risk of one's own life, which are normally displayed by a *positive* fraction of the overall community under examination. Fehr and Gächter (1999), by referring to sixteen different experimental studies, show that *reciprocally* and *selfishly* motivated people turn out to systematically coexist: in particular, they argue that in all scenarios[3] both types are present in non-negligible fractions, though the former seems to prevail. In the light of these observations, Hirshleifer & Martinez Coll's coexistence result as well as Palomino & Vega-Redondo's conclusion are quite interesting, especially if we think of the PD environment, as in such a large population framework, if we focused our attention on the adoption process of the 'classical' two pure strategies only (namely, 'cooperate' and 'defect'), coexistence between 'nice' and 'mean' strategies would be ruled out.

---

[1] Andreoni & Miller (1993) and Cooper et al. (1996) set up experiments where people play the Prisoner's Dilemma game sequentially with randomly changing opponents and find that while a minority of players act selfishly, the majority adopt non-selfish behaviors.

[2] Such an observation seems to be valid across countries and social contexts; on the contrary, in the light of empirical and experimental evidence, what appears to be strongly *culturally-specific* is the *relative frequency* with which the two types of behavior are observed.

[3] The experimental settings quoted by Fehr & Gächter (1999) include Prisoner's Dilemma, Investment Game, Public Goods Game and Trust Game. Fehr & Fischbacher (2002) argue that "during the last decade experimental economists have gathered overwhelming evidence that systematically refutes the self-interest hypothesis and suggests that a substantial fraction of the people exhibit social preferences, in particular, preferences for reciprocal fairness".

[4] Binmore (1994) points out that 'A society's pool of common knowledge - its culture, provides the informational input that individual citizens need to coordinate on *equilibria* in the games that people play. (…) An analyst ignorant of this data would not necessarily be able to predict the equilibrium on which members of the society would coordinate in a specific game. He might therefore categorize the equilibrium selection criteria that the society uses as arbitrary. However, the criteria will not seem arbitrary to those within the society under study'.

However, unlike these important contributions implicitly assume, it is far from obvious that individuals always interact with each other on the basis of a *clear* and *shared* perception of the overall social structure they are 'embedded' in. Such an assumption would require the presence of both a *very rich information set* on the part of the players and of a *very high degree of cultural proximity among them*, as it refers to a) agents' preferences, b) agents' ability to consistently and efficiently pursue their goals (i.e. their degree of 'rationality') and c) the outcomes associated to each couple of possible strategic choices. In other terms, the notion of 'common knowledge' turns out to be often a highly controversial and *context-dependent* one, as each individual's ability to 'frame' the social situation he is embedded in seems to be the effect of complex factors, active at both *natural* (e.g. cognitive and biological) and *cultural* (e.g. in terms of morality and social customs) level. For these reasons, individuals are likely to differ in the way they conceptualise the game they are about to play and such differences are likely to be specific to each sub-population. In other terms, we may say that in our framework 'common knowledge' about the payoff matrix has a sort of 'salience' (see Schelling 1960, Sahlins 1972) crucially dependent on the value-system characterizing the different types of players composing the overall community. In fact, it is reasonable to believe that the idea of salience not only regards *focal points* (sometimes described in terms of common knowledge about non-rational impulses; see e.g. Sugden 1991)[4] but, at a deeper level, the perception of the whole *payoff structure* of the game, depending on the locally prevailing social norms and cultural patterns.

On the basis of this approach, however, the relationship between rationality and salience should be somehow *reversed*, with respect to traditional game-theoretical frameworks where salience is a sort of 'second-best resource' agents rely on insofar as they fail to fruitfully coordinate their (individually rational) actions. To the contrary, in this paper we claim that agents, in the first place, tend to conceptualise the game they are about to play in a strongly culturally-dependent manner; then, at a second stage, rationality comes into the picture, inducing agents to choose the best strategies available on the basis of their information set. Clearly, insofar as salience regards the framing problem, *a fortiori* it can be claimed it concerns the more specific and conventional problem of focal points emergence. In the light of this, it is worth investigating the possibility of coexistence between 'nice' and 'mean' strategies in a different strategic context. In particular, we decide to focus on the following scenario: players have to choose one out of two strategies only (either 'cooperate' or 'not cooperate'); however, the possible outcomes of random pairwise

matchings are differently evaluated by single agents, i.e. individuals are heterogeneous in terms of their perception of the payoff matrix of the game they are involved in. As we pointed out above, agents are homogeneous only within specific sub-populations characterised by common socialisation patterns and salient social norms[5]: as far as the 'framing' problem of the initial multi-population community is concerned, then, *inter-group heterogeneity* corresponds to *intra-group homogeneity*. In other terms, we are still assuming *common knowledge* about the structure of the game to be played, but such a common knowledge is 'culturally-specific': each player's expectations about his opponent's behavior are systematically *biased* by his own reference culture and, therefore, confirmed only insofar as he happens to be matched with players of the same 'type'[6].

The idea of common knowledge we refer to recalls Lewis' definition (see Lewis 1969), concerned with justification and not with truth: what each person has reason to believe may be dependent on 'background information', which, we claim, is likely to depend in turn on his/her reference culture and social norms. This implies that agents' 'inductive standards' will be shared within each sub-population but will differ across them. Several experimental researches focusing on the effects of cultural background on game-theoretic behavior (see e.g. Smith & Bond 1993) confirm that a variable such as culture crucially affects the set of reference behavioral options individuals have in mind when playing standard games like PD. In particular, let us assume that the whole community consists of *three* sub-populations (types) of payoff-maximising individuals: in the first, everybody perceives the game matrix as the classical PD payoff configuration; in the second, agents believe an Assurance Game (AG) will have to be played, whereas in the third the payoff matrix is given by the Other Regarding game structure (OR). The purpose of our analysis is to consider the social dynamics taking place within such a complex environment. However, before introducing the evolutionary model, we want to suggest a motivationally-grounded interpretation of the differences in terms of 'framing' among the three sub-populations introduced above. In order to do this, let us consider the following Prisoner's Dilemma *payoff matrix*:

[5] It seems reasonable to assume that agents belonging to the same group have passed through similar socialization processes and therefore share common values and tend to conform to the same (population-specific) social norms: the majority of human customs and behaviors appears to be the consequence of complex processes of cultural evolution. Binmore (1994) remarks that 'A society's culture consists of more than the shared knowledge that we all belong to the same species. Vast amounts of historical data are enshrined in its customs and traditions'.

[6] 'A community of rational individuals is held together by the pool of common knowledge that I shall call its *culture*. The gossamer threads of shared knowledge and experience may seem flimsy bonds with which to hold a society together when compared with the iron shackles of duty and obligation postulated by traditional ethical theories. However, one must remember that the iron shackles of the traditionalists exist only in their imaginations, and even the most gossamer of real threads is more substantial than an iron shackle that is only imagined. Moreover, like Gulliver in Lilliput, we are bound by so many threads that even real shackles could fulfill their function with no greater efficiency' (Binmore 1994).

	Cooperate	Not cooperate
Cooperate	α, α	γ, δ
Not cooperate	δ, γ	β, β

where $\delta > \alpha > \beta > \gamma > 0$; $(\delta - \alpha) < (\beta - \gamma)$ and two rational players (A and B) are involved. We define agent A as 'altruist' when his utility is given by a weighted average of his own and agent B's payoff: $U_A = (1 - w)\,\Pi_A + w\,\Pi_B$, where Π_i ($i = A, B$) indicates i's payoff and w ($0 < w < 1$) represents A's (as well as B's) 'degree of altruism' toward her opponent. If both players are characterized by such a utility function, the *utility matrix* of the (symmetric) game becomes:

	Cooperate	Not cooperate
Cooperate	α, α	$(1 - w)\,\gamma + w\,\delta, (1 - w)\,\delta + w\,\gamma$
Not cooperate	$(1 - w)\,\delta + w\,\gamma, (1 - w)\,\gamma + w\,\delta$	β, β

When $0 < w < w_1 = (\delta - \alpha) / (\delta - \gamma)$, we fall into the classic PD game, whereas when $w_1 = (\delta - \alpha) / (\delta - \gamma) < w < w_2 = (\beta - \gamma) / (\delta - \gamma)$, the AG structure emerges; finally, when $w > w_2 = (\beta - \gamma) / (\delta - \gamma)$, we obtain the OR game. In other terms, the presence of three types of agents can be justified in terms of the perceived degree of altruism within one's reference sub-population: agents believe they are actually playing a PD, an AG, an OR game according to the level of w being *low* (equal to zero in the limit), *intermediate* or *high* (equal to one in the limit), respectively. The idea is that in a group where, say, pro-social values are traditionally rooted and widespread, it is reasonable to assume that each agent will both *act* on the basis of a *personal* pro-social attitude and *expect* his 'neighbors' to be driven by the same other-regarding motivational force: formally, this implies a *symmetric* game with $w > w_2$ will be played by such altruistically-driven agents. The same kind of considerations holds for less socially concerned individuals[7]: as Goldschmidt (1993) remarks, selective interaction

[7] For empirical evidence, see Ayres & Siegelman (1995) and Rapaport (1995) showing that market outcomes appear to systematically depend on the races of the parties involved; at experimental level, Weimann (1994) observes that in a repeated public good game framework American students turn out to be less cooperative than Germans, while Ockenfels & Weimann (1999) find that eastern Germans are far more selfish than western subjects.

[8] Regarding the interpretation in terms of culturally-specific motivational systems, it is important to clarify that we do not need to assume that OR players are actually driven by *genuinely altruistic* concerns: we can equivalently interpret their conceptualisation of the game as the effect of a sophisticated 'as if' calculating morality, letting them to implement the cooperative outcome and so to efficiently pursue their original selfish goals (see Sen 1974 for this intuition and Mueller 1986). On this view, people are assumed to choose the most efficient among alternative

tends to bring about common evaluations, as repeated social contact induces people to internalise others' positions and goals. Alternatively, the salience of utility matrices with low, intermediate or high values of w can be justified not in terms of individual motivational systems but as a consequence of *properly enforced social norms*: according to this explanation, players are still assumed to act on the basis of classic selfish preferences, but, at the same time, to be constrained by a culturally-specific set of pro-social norms prescribing how to behave in every feasible situation[8]. In particular, with reference to the above matrix, when $w < w_1$, it is 'as if' *no pro-social norms* were present or properly enforced in the group (D is the dominant strategy); when $w_1 < w < w_2$, then it is as if a *norm of reciprocity* or *conditional cooperation* were enforced and, finally, $w > w_2$ would imply a *norm of unconditional cooperation* (C is now the dominant strategy) is effectively at work.

The reason for choosing the specific payoff configurations under study (PD, AG and OR) is three-fold. First, they recall well-known and socially relevant scenarios (see Sen 1974). Second, they lend themselves to an analysis of social interaction taking place between individuals endowed with different degrees of *altruism* or, equivalently, between individuals conforming to different *social norms* (as we showed above). Third, *neither of them favours coexistence* - if taken separately from the others - between the two strategies under study. At methodological level, this means that if coexistence were to emerge in our scenario, such a result would provide a strong argument in favour of the main thesis defended here: by allowing for *heterogeneity* not simply in terms of individual strategies or motivational structures but in terms of group-specific 'game framing', we are able to provide a plausible explanation about *why* in many real social environments 'mean' and 'nice' strategies turn out to systematically coexist (though in different proportions) in the medium-long run. The plan of the remainder of the paper is as follows: Section 2 introduces the basic model; Section 3 develops the social dynamics; Section 4 concludes.

---

motivational structures, perceived as competing 'happiness technologies' (see Menicucci & Sacco 1997). An analogous explanation may be provided for AG players as well.

The model

The general framework is as follows: let us suppose that a *continuum* of agents belonging to a given community have to choose one out of H pure strategies $\{1,.....,H\}$ every time they interact with other individuals of the same community. Time is continuous. Individuals are distributed within M sub-populations $\{1,.....,M\}$, on the basis of their personal evaluation of the possible outcomes (in terms of pure strategies) of the random pairwise interactions. The M payoff configurations are assumed as exogenously given; more precisely, types that are initially present in the community may become extinct, but new types cannot be created. In this context, the outcome of an encounter between two individuals, let us call them I and II, is described by the pair (j,k), where the first and the second entry represent the pure strategies chosen by I and II respectively.

The adoption process of choices within the overall community is modeled by means of the so-called 'replicator equations' (see Taylor & Jonker 1978 and also Schlag (1994) and Björnerstedt & Weibull (1994) for some tentative micro-founded justifications of replicator dynamics): according to such equations, the most rewarding strategies survive and spread over within the community at the expense of the other. Such a selection mechanism affects both the width of each sub-population and the distribution of pure strategies within each sub-population. More precisely, we assume that social evolution not only operates at strategic level, but also at a deeper, *meta-behavioral level*, by selecting the most rewarding 'game framing' among PD, AG and OR. In other words, as far as each agent is concerned, 'game framing' is not to be interpreted here as an exogenous psychological or cultural feature or as an irreversible, one-shot decision (as if, for some reasons, agents had to stick forever to a given value-system and/or set of social norms), but as an *ongoing, endogenous process*, affected by both the sub-population type he/she belongs to and the reward he/she gets by his/her choice. The idea is then to test how the three different sub-populations (types) initially present within the community are *evolutionarily robust* in the sense of being able to attract an increasing number of adherents at the expense of the alternative ones. We further assume that the payoffs corresponding to each pair (j,k) depend on the population to which individuals belong. In particular, we will focus on two very different cases:

(a) In the first, the payoff of player I (II) related to the event (j,k) depends on the population he belongs to and not on the population of the opponent player II (I). In this case, the payoff of player I, belonging to population i and related to the event (j,k), is expressed by the symbol a_{ijk}, where $i=1,...,M$ and $j,k=1,...,H$. Notice that if, given two populations i^* and i^{**}, $a_{i^*jk} > a_{i^{**}jk}$ holds whatever (j,k) is, then "to belong to type i^*" is always more rewarding than "to belong to

type i^{**} ". In such a case, social dynamics turns out to be very simple: type i^{**} becomes extinct. However, we shall mainly deal with the more general (and interesting) case in which such a strict payoff dominance does not hold;

(b) In the second case, which includes the first as a particular case, we assume that the payoff of I (II) related to the result (*j,k*) *also* depends on the *population* to which the opponent player II (I) belongs[9]. I's payoffs are expressed by the symbol a_{ijkl}, where the index l=l,....,M represents the population of the opponent player. The specific meaning of this assumption will be subsequently clarified.

The dynamics under study can be interpreted as follows: the structure of the community outlines a preference ordering which is not based on outcomes (*j,k*), but on more complex outcomes (*i,j,k,l*): "to be an individual of type *i*, playing the pure strategy *j* on the occasion of an encounter with an individual of type *l*, playing the pure strategy *k*". In the following sections we will analyze these two cases separately. A rapidly growing literature considers payoffs as not univocally determined by the "material" outcomes of the strategic interaction taking place between players. Payoffs are more and more considered as the result of the interaction between "material" and "immaterial" components[10], related to the social and cultural environment in which individuals act; see, e.g. Wildavsky (1992) and Fehr & Fischbacher (2002); see also Antoci et al. (1998), which contains a wide review of such literature. In particular, this study builds on the work of Sacco & Zamagni (1996) and has various connections with it. Both contributions analyse heterogeneous communities and social dynamics based on the selection of the most rewarding strategies.

Nevertheless, there are some substantial differences between the two papers. While Sacco & Zamagni assume that individuals have payoffs of the above described case (a), they do not consider case (b). Here we assume that each individual can *only* recognize *ex post* the sub-population type his opponent belongs to and the pure strategy played by him. In contrast, Sacco & Zamagni postulate that individuals are able to recognize *ex ante* the opponent player type and thus Nash equilibria are played in each matching. Consequently, social dynamics runs over the proportions of types only. Finally, their paper does not highlight phenomena of coexistence among the different types of players with which they deal. In this work, we shall indicate with

[9] Referring to Granovetter's (1985) fundamental work, Sacco & Zamagni (1996) remind us that 'Individual behaviours are *embedded* in a preexisting network of social relations which cannot be thought as a mere constraint; rather, they are one of the driving forces that prompt individual goals and motivations'.

[10]The plausibility of this payoff structure, whose exact meaning will be better explained subsequently, is supported by several contributions. Smith-Lovin (1993) claims that 'all people are emotional in predictable ways; what emotions they feel after a given transaction depends on the culture in which they are embedded, the character of the relationship between partners, and the type of exchange that occurs between them'.

the term "action" the pair (*i,j*) where $i = 1,.....,M$ and $j = 1,.....,H$ respectively indicate the population and the pure strategy chosen by an individual. In this context, a_{ijk} represents the payoff of an individual of population *i*, choosing pure strategy *j* when the opponent player chooses pure strategy *k*, $k = 1,.....,H$.

Independence of the opponent's type

Let us examine, in the first place, the 'conventional' case in which each player's payoff does *not* depend on the population of the opponent player but only on the pure strategy followed by the latter. We assume a community with a very large number of individuals. Let x_{ji} be the proportion (w.r.t. the whole community) of individuals belonging to population *i* and following pure strategy *j*; thus:

$$x_j \equiv \sum_{i=1}^{M} x_{ji}$$

represents the proportion of the community playing pure strategy *j*. Each individual knows the opponent player's type ex post only, i.e. after both players have played their pure strategies. Therefore, individuals are not able to play best responses. The expected payoff Y_{ji} of the action (*i,j*) is:

$$Y_{ji} \equiv \sum_{k=1}^{H} \sum_{l=1}^{M} x_{lk} a_{ijk} \qquad (1)$$

where $i = 1,.....,M$ and $j = 1,.....,H$. Notice that the proportions x_{ji} and x_j can also be interpreted as the probabilities that the opponent player respectively follows action (*i,j*) and pure strategy *j*. The mean payoff Y of the community is:

$$Y \equiv \sum_{j=1}^{H} \sum_{i=1}^{M} x_{ji} Y_{ji} \qquad (2)$$

Dependence on the opponent's type

In this context, the payoffs of individuals of sub-population *i* depend not only on the pure strategy played by the opponent player, but also on the sub-population to which the opponent player belongs, i.e. on the action he chooses[11]. The rationale of this methodological choice is as follows: we assume that a given outcome (i.e. pair of strategies) can bring about different values in terms of overall individual payoff (i.e. 'utility') according to how each agent evaluates his opponent's 'game framing', which in turn, as we previously clarified, crucially depends on the specific social norms and cultural patterns characterizing each sub-population. In particular, it seems reasonable to assume that for an AG player cooperating when the opponent defects determines a *lower payoff* if the opponent is a PD agent rather than an AG agent or an OR agent, as the AG player knows that PD players are presumably selfish (or, equivalently, act as if they were driven by basically self-interested social norms) and, unlike OR agents, tend to *exploit* their opponents[12]. More precisely, we now consider the payoff of player of type *i* playing pure strategy *j* when matched with a player of type *s* playing pure strategy *k*: a_{ijks}. As above, x_{ji} represents the proportion of individuals belonging to population *i* playing pure strategy *j*, and the expected payoff of playing *j* by an individual of the type *i* is:

$$Y_{ji} \equiv \sum_{k=1}^{H} \sum_{l=1}^{M} x_{lk} a_{ijkl} \tag{3}$$

where $i = 1,.....,M$ and $j = 1,.....,H$; and the mean payoff of the community is:

$$Y \equiv \sum_{j=1}^{H} \sum_{i=1}^{M} x_{ji} Y_{ji} \tag{4}$$

[11] Such interaction could, even radically, modify the 'purely material' payoff structure and, consequently, the choices determined by them, as we showed in section 1 by illustrating how different levels of the 'degree of altruism' w can lead to different payoff configurations, such as PD, AG and OR (see Taylor 1987 for a rigorous analysis).

[12] Similarly, Banerjee & Weibull (1995) set up a 'discriminating players' model where agents are able to identify their opponents' type and to consequently act on the basis of an 'opponent-sensitive' logic of play.

The analysis of this simple strategic and social context will allow us to show that:

(1) Even though these specific payoff configurations (if taken separately) do not generate coexistence-favouring social dynamics, coexistence may take place in the heterogeneous community under examination.

(2) Even in the simplest strategic context (two pure strategies), social dynamics may turn out to be quite complex, unlike the case of a homogeneous community where individuals have to choose between two pure strategies only.

(3) By taking into consideration only initial aggregate proportions of agents choosing between 'cooperate' and 'not cooperate' (i.e. leaving out of consideration the types of individuals initially adopting each strategy), rather misleading predictions may be obtained; that is, final results of social dynamics may be radically different for very similar initial aggregate proportions.

Social dynamics

As we anticipated above, the selection mechanism of actions is modelled by means of the so-called 'replicator equations' (Taylor & Jonker 1978):

$$\dot{x}_{ji} = x_{ji}\left(Y_{ji} - Y\right) \tag{5}$$

where $i = 1,.....,M$ and $j = 1,.....,H$. Following Weibull (1995), we can obtain (5) as follows. Assume that the number of individuals in the community is very large; let $p_{ji}(t) \geq 0$ be the number of individuals choosing action (j,i) and let $p(t) \equiv \sum_{ji} p_{ji}(t)$ be the total number of individuals in the community; thus $x_{ji}(t) = p_{ji}(t) / p(t)$. Let us assume that all individuals have a background fitness, measured as the number of offsprings per time unit $\beta \geq 0$ and a death rate $\delta \geq 0$ that are independent of their performance in the game under study. Augmenting this "biological" replicator process by the corrective factor Y_{ji}, population dynamics can be represented as follows:

$$\dot{p}_{ji} = p_{ji}(\beta + Y_{ji} - \delta) \tag{5'}$$

It is easy to show (see Weibull 1995, pp. 72-73) that dynamics (5') imply dynamics (5) for population shares. Dynamics (5) are defined on the invariant simplex:

$$\Delta = \left\{ \mathrm{x} \in \Re^{MH}, \sum_{j=1}^{H} \sum_{i=1}^{M} x_{ji} = 1, x_{ji} \geq 0 \right\}.$$

Notice that the states of the community in which all individuals choose the same action are fixed points under dynamics (5). The other fixed points are the states of the community in which the actions representing a positive proportion of the community yield the same expected payoffs, i.e. there exists a constant Y^* such that:

$$Y_{ji} = Y^* \tag{6}$$

for every action (*j,i*) such that $x_{ji} > 0$.

Notice also that if payoffs do not depend on the opponent player type, condition (6) can be written:

$$Y_{ji} = \sum_{k=1}^{H} x_k a_{ijk} = Y^* \tag{7}$$

where x_k is the proportion in the community of players playing pure strategy *k*.

System (7) is a linear system where the number of equations is equal to the number of actions (*j,i*) by which $x_{ji} > 0$ and the number of unknowns is equal to the number of pure strategies *k* such that $x_k > 0$. Therefore, such a system does not generically admit solution if the number of actions is greater than the number of pure strategies that are played in the community. This implies that the fixed points we may generically observe are those with *H* actions at most, where *H* is the number of pure strategies that are initially present in the community. It also follows that the *maximum number of sub-populations* that may coexist at a fixed point is equal to the *number of available pure strategies*. This means that the degree of complexity of the social structure is closely related to the number of pure strategies available.

Through this analytical framework, we will analyse the process of cultural evolution taking place within a large community in which there are *two* pure strategies only ('cooperate' (C) and 'not cooperate' (NC)), and *three* sub-populations in total. In population 1 individuals have

Prisoner's Dilemma (PD) payoffs. Let us recall that if we have two players, I and II, the four possible outcomes of the PD game (from the point of view of player I) follow the order:

$$(NC,C) \succ (C,C) \succ (NC,NC) \succ (C,NC)$$

where the first entry of each pair represents the strategy chosen by I and $\succ$ indicates *strict* preference. If we assign indexes 1 and 2 to strategies NC and C respectively, PD payoffs satisfy the following inequalities:

$$a_{112} > a_{122} > a_{111} > a_{121}\,.$$

In population 2, individuals have *Assurance Game* (AG) payoffs (see Sen 1967), i.e.:

$$(C,C) \succ (NC,C) \succ (NC,NC) \succ (C,NC)$$

and consequently:

$$a_{222} > a_{212} > a_{211} > a_{221}\,.$$

In this game, players show both *positive reciprocity* (being kind to those who have been kind to them) and *negative reciprocity* (by retaliating if they have been hurt), that is their propensity to cooperate is *conditional* on their opponent's behavior. Fehr & Gächter (1999) and Fehr & Fischbacher (2002) show that there is strong experimental and empirical evidence that agents exhibit *both types of reciprocation* and that this behavior occurs even in one-shot encounters between strangers and when retaliation is costly and yields neither present nor future material rewards. For surveys of experimental results documenting the frequency of reciprocity in Ultimatum Bargaining Games, Gift-Exchange Games and Trust Games, see e.g. Güth et al. (1982), Camerer & Thaler (1995), Fehr et al. (1993) and Roth (1995).

Finally, in population 3 individuals have *Other Regarding* (OR) payoffs, i.e.:

$$(C,C) \succ (C,NC) \succ (NC,C) \succ (NC,NC)$$

and consequently:

$$a_{322} > a_{321} > a_{312} > a_{311}\,.$$

In PD and OR populations, the strategies NC and C respectively (*strictly*) *dominate* the alternative strategy. Therefore, without loss of generality (see e.g. Weibull 1995), we can analyse dynamics (5) by assuming that *no player* in these populations chooses the *dominated* strategies. We shall indicate by NC_{PD} and NC_{AG} the actions "to be a PD individual playing strategy NC" and "to be an AG individual playing strategy NC", respectively; and by C_{OR} and C_{AG} the actions "to be an OR individual playing strategy C" and " to be an AG individual playing strategy C", respectively.

Bistable dynamics

Let us first analyse a case in which payoffs do *not* depend on the opponent's type and coexistence is ruled out. Let us consider the payoff structure given by the following matrix (from the point of view of the row player):

INSERT TABLE 1 ABOUT HERE

Dynamics (5) run over four variables, x_{11}, x_{21}, x_{22} and x_{32}, representing the proportions of individuals following the actions NC_{PD}, NC_{AG}, C_{OR} and C_{AG} respectively:

$$\begin{aligned}
\dot{x}_{11} &= x_{11}\left[(\mathrm{Ax})_1 - {}^t\mathrm{x}\cdot\mathrm{Ax}\right] \\
\dot{x}_{21} &= x_{21}\left[(\mathrm{Ax})_2 - {}^t\mathrm{x}\cdot\mathrm{Ax}\right] \\
\dot{x}_{22} &= x_{22}\left[(\mathrm{Ax})_3 - {}^t\mathrm{x}\cdot\mathrm{Ax}\right] \\
\dot{x}_{32} &= x_{32}\left[(\mathrm{Ax})_4 - {}^t\mathrm{x}\cdot\mathrm{Ax}\right]
\end{aligned} \tag{8}$$

where ${}^t\mathrm{x} \equiv (x_{11}, x_{21}, x_{22}, x_{32})$, A is the above payoff matrix

$$\begin{bmatrix} 4 & 4 & 8 & 8 \\ 5 & 5 & 7 & 7 \\ 3 & 3 & 9 & 9 \\ 2 & 2 & 10 & 10 \end{bmatrix}$$

and $(A\mathrm{x})_r$ is the r-th component of the vector $A\mathrm{x}$. In this specific case, the state space of dynamics (8) is the (3-dimensional) simplex $\Delta = \{\mathrm{x} \in \Re^4 : \mathrm{x} \geq 0 \text{ and } x_{11} + x_{211} + x_{22} + x_{32} = 1\}$. By the illustrative device adopted in Hirshleifer & Martinez Coll (1991), we can represent the edges of Δ (i.e. the boundary of the simplex in which at least one action is extinct) in the plane (see figure 1). Thus the simplex Δ can be imagined as based on the triangle $NC_{PD} - NC_{AG} - C_{AG}$, while C_{OR} is the upper vertex, that in which all actions are extinct except for C_{OR} (by drawing the edges in the 3-dimensional euclidean space, all the C_{OR} vertices in figure 1 will come together).

INSERT FIGURE 1 ABOUT HERE

In figure 1, the dynamics on the edges are obtained by means of Bomze's (1983) classification technology for 2-dimensional replicator equations. Following Bomze's symbology, a dotted line represents a line of fixed points (pointwise fixed), a full dot • represents a fixed point which is locally attractive, whereas saddle points are indicated by their insets and outsets (stable and unstable manifolds, respectively). Only some representative trajectories are sketched. From figure 1, we can see that social dynamics bring about a "bistable dynamics", i.e. the only attractive fixed points are the vertices C_{OR} and NC_{AG} and their attraction basins are separated by a 2-dimensional (repulsive) pointwise fixed set in the interior of Δ, whose intersection with the edges is given by the pointwise fixed lines shown in figure 1. If large enough proportions of individuals choosing action C_{OR} (NC_{AG}) are initially present, then all actions except for C_{OR} (respectively, NC_{AG}) become extinct. The social structure that eventually emerges is very simple: a single population playing only one pure strategy is present[13]. In the following example, we consider dynamics starting from a payoff structure which strongly favours coexistence.

[13] Sahlins (1972) refers to a similar selective attitude describing human societies where the same agent consistently displays a cooperative attitude toward people he feels 'close' to as well as a payoff-maximizing or even hostile attitude towards people he perceives as 'strangers'.

Coexistence-favouring payoffs

Let us consider the following payoff matrix, where we still assume that players' payoffs do not depend on the opponent's type:

INSERT TABLE 2 ABOUT HERE

In this example, the highest payoff level is reached by a PD individual when matched with an individual playing the pure strategy C. On the other hand, the payoff of a PD individual is very low when he is matched with an individual playing NC. This rules out the emergence of a homogeneous community where only the PD sub-population exists. On the contrary, OR individuals assign a relatively high payoff to the outcome (C, NC), while their payoff associated to the outcome (C,C) is relatively low. In this case, as in the previous one, we cannot expect a homogenous OR-type community to emerge, as such a community would turn out to be extremely vulnerable with respect to a population of PD players. The other entries of this coexistence-favouring payoff matrix can be interpreted in a totally analogous way. The phase portrait at the edges of the simplex Δ is represented in figure 2.

INSERT FIGURE 2 ABOUT HERE

It is easy to verify that there are no fixed points in which more than three actions coexist. Thus, by a well-known result (see Weibull 1995), under dynamics (8) trajectories always approach the edges represented in figure 2. In this figure, we can notice that, starting from "almost all" the initial distributions of actions in the community, the social dynamics reach a fixed point in which AG and OR-type sub-populations coexist playing NC_{AG} and C_{OR} respectively. In the last example of this section, we focus on a mixed case where a configuration of bistable dynamics is merged with a different one where coexistence emerges. Let us examine the following payoff matrix:

INSERT TABLE 3 ABOUT HERE

This payoff structure is characterized by the fact that both C_{OR} and NC_{AG} individuals perform very well with opponent players of the same type. Therefore, the vertices C_{OR} and NC_{AG} are both locally attractive (see figure 3). However, if the proportions of NC_{PD} and C_{AG} individuals are large enough, a two-population community in which only these two types coexist may

emerge. As above, it is easy to verify that no fixed points exist with more than two actions. Thus, figure 3 represents the "limit" dynamics of (8). We shall discuss this case further in the last section, when some features of aggregate dynamics will be analysed.

INSERT FIGURE 3 ABOUT HERE

Let us now turn our attention to the following payoff matrix, where payoffs depend not only on the strategic *choices* of the two players but also on the opponent player's *type*:

INSERT TABLE 4 ABOUT HERE

Notice that we have introduced here two parameters, α and β, where $0<\alpha<1$ and $\beta>0$, through which it is possible to account for the qualitative modifications of social dynamics. Such a matrix shows that whereas PD player's payoffs are completely independent of the population his opponent belongs to, both AG and OR players' payoffs crucially depend on their opponent's type. In particular, an AG player gets a lower payoff when cooperating with a defecting PD rather than with a defecting AG, given his awareness of the deeply *selfish* nature of PD agents (or, at least, of their unambiguously anti-cooperative behavior). Analogously, AG players can be expected to be 'happier' when defecting with a defecting PD rather than with a defecting AG. Further, AG-type agents get a higher payoff when cooperating with a cooperating OR rather than with a cooperating AG. The rationale behind these payoff differences can be explained as follows: OR players are perceived as *more trustworthy* agents as they tend to *cooperate unconditionally*, that is to never defect regardless of their opponent's behavior. In other terms, we can plausibly imagine a sort of 'moral ranking' among the three types, according to which – as far as the opponent's choices are concerned - OR behavior is preferable to AG behavior which in turn can be considered as morally superior to PD behavior. Therefore, as it is immediate to observe from the above matrix, OR players prefer to cooperate with a defecting AG agent rather than with a defecting PD-type agent and with a cooperating OR agent rather than with a cooperating AG-type. As anticipated above, when payoffs depend on the opponent player's type, far more complex social structures are likely to emerge; in particular, fixed points with more than two actions are not ruled out in generic cases.

In our example, it is easy to verify that a fixed point P, in which all the actions are present, exists if and only if $\beta < 1/2$; further, it is always locally attractive (see the mathematical appendix) and its coordinates are:

$$(x^*_{11}, x^*_{21}, x^*_{22}, x^*_{32}) = \frac{1}{16 - 7\alpha\beta - 4\beta}(6 - 3\alpha\beta, 1 - 2\beta, 5 + \alpha - 4\alpha\beta - 2\beta, 4 - \alpha)$$

Thus, a social configuration with three sub-populations playing four actions can be locally attractive under dynamics (8). Notice that, in such a configuration, both C_{AG} and NC_{AG} individuals coexist, whereas such a coexistence pattern is ruled out in a community where only an AG population is initially present. The dynamics driven by the above payoff matrix is interesting also because, by changing parameter values, a relatively wide "zoology" of cases can emerge. In the following figures, we only sketch the "representative" ones. Since fixed points with more than two actions may exist under such a payoff matrix, their stability cannot be checked by reference to Bomze's classification only; it is also necessary to use the standard procedure of local analyses (see the mathematical appendix). We consider four cases:

Case (a): For $\alpha = 1/4$ and $\beta \geq 4$, the fixed point P does not exist; thus trajectories always approach the edges of Δ. The dynamics on the edges is given in figure 4.

INSERT FIGURE 4 ABOUT HERE

We can observe that "almost all" the trajectories approach the vertex C_{OR}. Notice that, in this case, the action C_{OR} performs better against itself than the action NC_{PD} against C_{OR}. Notice also that, if the OR population is extinct (see triangle $NC_{PD} - NC_{AG} - C_{AG}$) we have a fixed point surrounded by closed trajectories. However, it is easy to see that such trajectories become repulsive when the OR population is introduced into the community (see the mathematical appendix).

Case (b): For $\alpha > 1/4$ and $3 < \beta < 4$, the fixed point P does not exist and the dynamics on the edges is shown in figure 5.

INSERT FIGURE 5 ABOUT HERE

In this case, "almost all" trajectories approach a fixed point in which both C_{OR} and NC_{PD} coexist. In the mathematical appendix we show that the fixed point in the interior of the triangle $NC_{PD} - NC_{AG} - C_{AG}$ (which is attractive on the edges) is a saddle point, i.e. it is unstable.

Case (c): For $\alpha = 1/4$ and $1 \le \beta < 3/2$, the fixed point P does not exist and the dynamics on the edges are shown in figure 6.

INSERT FIGURE 6 ABOUT HERE

In this case, the fixed point in which both C_{OR} and NC_{PD} coexist becomes unstable; the fixed point in the interior of the triangle $NC_{PD} - NC_{AG} - C_{AG}$ remains repulsive, while almost all the trajectories are attracted by the fixed point in the interior of the triangle $NC_{PD} - C_{AG} - C_{OR}$. All the sub-populations coexist in this fixed point and, as above, social dynamics reach a fixed point in which both strategies C and NC coexist.

Case (d): For $\alpha = 1/4$ and $1/4 \le \beta < 1/2$, the locally attractive fixed point P exists; at such a point, no population becomes extinct and AG individuals play both C and NC. The dynamics on the edges (shown in figure 7) is analogous to that of figure 6; however, in this case, the fixed point in the interior of the triangle $NC_{PD} - C_{AG} - C_{OR}$ becomes a saddle point (see the mathematical appendix).

The local attractivity of the fixed point P does not imply its global attractivity. In fact, in the interior of the state space Δ, other attractors may exist. However, even if in this case P may be a global attractor, it is surely possible to construct *ad hoc* payoff matrices according to which dynamics (8) have a strange attractor in the interior of the state space Δ. In such a case, OR, AG and PD sub-populations in this community coexist, all playing pure strategies C and NC, although the dynamics never reach a fixed point. Furthermore, the outcome of social dynamics can be critically dependent on initial distributions of actions in the community; in such a case, social dynamics is unpredictable, at least from a deterministic point of view. To build these ad hoc matrices, see Schnabl et al. (1991); in particular, see matrices (7)-(9) of their paper.

INSERT FIGURE 7 ABOUT HERE

Concluding remarks

In order to stress the importance of our results, let us recall that for symmetric two-player games with two pure strategies (e.g. NC and C) played in a homogeneous community:

	NC	C
NC	a_{11}	a_{12}
C	a_{21}	a_{22}

we can only have four (generic) dynamic regimes under replicator dynamics (see Weibull 1995, pp.74-76):

(i) For $a_{11} > a_{21}, a_{12} > a_{22}$ the pure strategy NC (strictly) dominates C; in this case, the share of individuals choosing NC approaches the value 1 when time goes to infinity.

(ii) For $a_{11} < a_{21}, a_{12} < a_{22}$ the opposite case holds.

(iii) For $a_{11} > a_{21}, a_{12} < a_{22}$ both the pure population states, in which all the individuals respectively play NC or C, are locally attractive fixed points; their attraction basins are separated by a repulsive fixed point in which both strategies are played.

(iv) For $a_{11} < a_{21}, a_{12} > a_{22}$ there is a globally attractive fixed point where both strategies coexist.

In such a context (two pure strategies and a homogeneous community), payoff configuration (iv) only admits coexistence between NC and C. According to the others, we expect to see individuals playing only one strategy after transient dynamics. Therefore, the coexistence of strategies can be explained only through very restrictive assumptions over individual payoffs. This prediction is rather unrealistic in a world where we generally observe coexistence between "nice" and "mean" strategies. Hirshleifer & Martinez Coll (1991) assume homogeneous interactions but add to the set of pure strategies related to the games PD and CG two "reactive" strategies, such as 'tit for tat' (a "nice" strategy) and 'bully' (a "mean" strategy). The games PD and CG are played separately; i.e. they first consider replicator dynamics under a Prisoner's Dilemma environment and then study dynamics for the Chicken game. They show that, in this context, dynamics exhibit very interesting features. More specifically, they show that dynamics can be substantially more complex than the dynamics of regimes (i)-(iv) and that we can expect, under both payoff environments, coexistence between "nice" and "mean" strategies. The complexity of dynamics studied by Hirshleifer & Martinez Coll is a direct consequence of the assumption that players are able to play reactive strategies. Unlike that important contribution,

let us recall that in our paper we proceed by postulating that all individuals in the community play two (non-reactive) pure strategies only but, at the same time, they are *heterogeneous* w.r.t. their way of *framing* the game, which is *culturally-specific* (i.e. specific to each sub-population). Furthermore, we obtain coexistence results even when each individual has payoffs that do not favour coexistence, i.e. a type (i), a type (ii) or a type (iii) individual payoffs configuration.

APPENDIX

The dynamics under system (8) for payoffs which do not depend on the opponent player, can be analysed by Bomze's results (see Bomze 1983). When payoffs depend on the opponent player, we obtain fixed points in which more than two actions coexist. In these cases, to analyse local stability, it is necessary to linearize system (8) around these fixed points. To this end, we use the well-known correspondence between replicator equations and Lotka-Volterra systems (see Hofbauer & Sigmund 1988). In particular, in this case, we have that the transformation

$$T:(y,z,w)\to(x_{11},x_{21},x_{22},x_{32})=$$

$$=\left(\frac{1}{1+y+z+w},\frac{y}{1+y+z+w},\frac{z}{1+y+z+w},\frac{w}{1+y+z+w}\right).$$

maps the trajectories under Lotka-Volterra equations:

$$\begin{aligned}\dot{y}&=y[4+3y-3z-3w]\\ \dot{z}&=z[2+(2+\alpha)y-2z-w]\\ \dot{w}&=w[5+6y-4z+(\beta-4)w]\end{aligned}\qquad(9)$$

onto those of replicator equations (8).

The inverse transformation of T is:

$$T^{-1}:(x_{11},x_{21},x_{22},x_{32})\to(y,z,w)=\left(\frac{x_{21}}{x_{11}},\frac{x_{22}}{x_{11}},\frac{x_{32}}{x_{11}}\right)$$

System (9) has a unique fixed point at which y, z and w are strictly positive if and only if $\beta < 1/2$. In this case, the fixed point has coordinates:

$$\left(y^*, z^*, w^*\right) = \frac{1}{3(\alpha\beta - 2)}\left(2\beta - 1, 4\alpha\beta + 2\beta - \alpha - 5, \alpha - 4\right)$$

Otherwise, it has no fixed points where all the actions are simultaneously present in the community. In the original coordinates, the above fixed point becomes:

$$(x^*_{11}, x^*_{21}, x^*_{22}, x^*_{32}) = \frac{1}{16 - 7\alpha\beta - 4\beta}\left(6 - 3\alpha\beta, 1 - 2\beta, 5 + \alpha - 4\alpha\beta - 2\beta, 4 - \alpha\right)$$

The Jacobian matrix *J* of (9), evaluated at $\left(y^*, z^*, w^*\right)$, has entries J_{ij}:

$$J_{11} = 3y^*, J_{12} = -3y^*, J_{13} = --3y^*$$

$$J_{21} = (2 + \alpha)z^*, J_{22} = -2z^*, J_{23} = -z^*$$

$$J_{31} = 6w^*, J_{32} = -4w^*, J_{33} = (\beta - 4)w^*$$

By Routh-Hurwitz criterion (see e.g. Beavis & Dobbs 1989, p. 134), a necessary and sufficient condition for local asymptotic stability of $\left(y^*, z^*, w^*\right)$ is that $TrJ < 0$, $DetJ < 0$ and $Det\widetilde{J} < 0$ where:

$$\widetilde{J} \equiv \begin{bmatrix} J_{11} + J_{22} & J_{23} & -J_{13} \\ J_{32} & J_{11} + J_{33} & J_{12} \\ -J_{31} & J_{21} & J_{22} + J_{33} \end{bmatrix}$$

It is easy to see that this system meets these conditions.

A fixed point with $y = 0$ and $z, w > 0$, i.e. a fixed point in which only $x_{21} = 0$, exists if and only if $\beta < 3/2$. In this case, it has the following coordinates:

$$(\bar{y},\bar{z},\bar{w}) = \left(0, \frac{1}{2-\beta}, \frac{3-2\beta}{2(2-\beta)}\right)$$

The Jacobian matrix evaluated at $(\bar{y},\bar{z},\bar{w})$ is:

$$\begin{bmatrix} 4-3\bar{z}-3\bar{w} & 0 & 0 \\ (2+\alpha)\bar{z} & -2\bar{z} & -\bar{z} \\ 6\bar{w} & -4\bar{w} & (\beta-4)\bar{w} \end{bmatrix}$$

Notice that the eigenvalue in the direction of the interior of the simplex Δ:

$$4-3\bar{z}-3\bar{w} = \frac{1-2\beta}{2(2-\beta)}$$

is strictly positive if and only if $\beta < 1/2$, i.e. when the interior fixed point exists in the simplex. We have a fixed point with $z = 0$ and $y, w > 0$, corresponding to the fixed point in which only $x_{22} = 0$, if and only if $\beta < 1/4$. In this case, it has coordinates:

$$(\hat{y},\hat{z},\hat{w}) = \left(\frac{1-4\beta}{3(2+\beta)}, 0, \frac{3}{2+\beta}\right)$$

and the relative Jacobian matrix:

$$\begin{bmatrix} 3\hat{y} & -3\hat{y} & -3\hat{y} \\ 0 & 2+(2+\alpha)\hat{y}-\hat{w} & 0 \\ 6\hat{w} & -4\hat{w} & (\beta-4)\hat{w} \end{bmatrix}$$

has the following strictly positive eigenvalue in the direction of the interior of the simplex:

$$2+(2+\alpha)\hat{y}-\hat{w} = \frac{5+\alpha-2\beta-4\alpha\beta}{3(2+\beta)}.$$

We always have a fixed point with $w = 0$ and $y, z > 0$ (i.e. with only $x_{32} = 0$) and it has the following coordinates:

$$(\hat{y},\hat{z},\hat{w}) = \left(\frac{2}{3\alpha},\frac{4\alpha+2}{3\alpha},0\right)$$

with the Jacobian matrix

$$\begin{bmatrix} 3\hat{y} & -3\hat{y} & -3\hat{y} \\ (2+\alpha)\hat{z} & -2\hat{z} & -\hat{z} \\ 0 & 0 & 5+6\hat{y}-4\hat{z} \end{bmatrix}$$

We can see that it has the following strictly positive eigenvalue in the direction of the interior of the simplex:

$$5+6\hat{y}-4\hat{z} = \frac{4-\alpha}{3\alpha}.$$

References cited

Andreoni, James & John Miller, 1993. Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence. Economic Journal 103: 570-585.

Antoci, Angelo & Pier Luigi Sacco, 1995. A public contracting evolutionary game with corruption. Journal of Economics 61: 89-122.

Antoci Angelo, Pier Luigi Sacco & Stefano Zamagni, 2000. The ecology of altruistic motivations in triadic social environments. In L. A. Gérard-Varet, S. C. Kolm, J. Mercier Ythier (eds.) The Economics of Reciprocity, Giving and Altruism, IEA Conference volume, Macmillan, London**.**

Ayres, Ian. & Peter Siegelman, 1995. Race and gender discrimination in bargaining for a new car. American Economic Review 85 (3): 304-319.

Banerjee, Abhijit Vinayak & Jörgen Weibull, 1995. Evolutionary selection and rational behaviour. In A. Kirman, M. Salmon (eds.) Learning and rationality in economics, Basil Blackwell, Oxford.

Beavis, Brian & Ian Dobbs, 1989. Optimization and stability theory for economic analysis. Cambridge University Press, Cambridge.

Binmore, Ken. 1994. Game Theory and the Social Contract. Volume I: Playing Fair. The MIT Press, Cambridge (MA).

Binmore, Ken & Larry Samuelson, 1992. Evolutionary Stability in Repeated Games Played by Finite Automata. Journal of Economic Theory 57: 278-305.

Björnerstedt, Jonas & Jörgen Weibull, 1994. Nash equilibrium and evolution by imitation, mimeo, Delta, Paris.

Bomze, Immanuel. 1983. Lotka-Volterra Equation and Replicator Dynamics: a Two-Dimensional Classification. Biological Cybernetics 48: 201-211.

Bomze, Immanuel. 1986. Non-cooperative two-person games in biology: a classification. International Journal of Game Theory 15:31-57.

Camerer, Colin & Richard Thaler, 1995. Ultimatums, Dictators, and Manners. Journal of Economic Perspectives 9: 209-219.

Cooper, Russell, Douglas DeJong, Robert Forsythe & Thomas Ross, 1996. Cooperation without Reputation: Experimental Evidence from Prisoner's Dilemma Games. Games and Economic Behavior 12: 187-218.

Ellison, Glenn. 1994. Cooperation in the prisoner's dilemma with anonymous random matching. Review of Economic Studies 61: 567-588.

Eshel, Ilan, Sansone, Emilia & Avner Shaked, 1999. The emergence of kinship behavior in structured populations of unrelated individuals. International Journal of Game Theory 28: 447-463.

Fehr, Ernst & Urs Fischbacher, 2002. Why social preferences matter – The impact of non-selfish motives on competition, cooperation and incentives. The Economic Journal 112: 1-33.

Fehr, Ernst & Simon Gächter, 1999. Reciprocal Fairness, Heterogeneity, and Institutions. Paper presented at the AEA Meeting in New York, Jan. 3-5 1999.

Fehr, Ernst, Kirchsteiger, Georg & Arno Riedl, 1993. Does fairness prevent market clearing? An experimental investigation. Quarterly Journal of Economics 108: 437-46

Goldschmidt, Walter. 1993. On the Relationship Between Biology and Anthropology. Man 28: 341-359.

Granovetter, Mark. 1985. Economic Action and Social Structure: The Problem of Embeddedness. American Journal of Sociology 91: 481-510.

Güth, Werner, Rolf Schmittberger & Bernd Schwarze, 1982. An experimental analysis of ultimatum bargaining. Journal of Economic Behavior and Organization 3: 367-388.

Hirshleifer, Jack & Juan Carlos Martinez Coll, 1991. The limits of reciprocity. Rationality and Society 3: 35-64.

Hofbauer, Josef & Karl Sygmund, 1988. The Theory of Evolution and Dynamical Systems. Cambridge University Press, London.

Kandori, Michihiro. 1992. Social Norms and Community Enforcement. Review of Economic Studies 59: 63-80.

Karandikar, Rajeeva, Dilip Mookherjee, Debraj Ray & Fernando Vega-Redondo, 1998. Evolving aspirations and cooperation. Journal of Economic Theory 80: 292-331.

Lewis, David. 1969. Convention. A Philosophical Study. Harvard University Press, Cambridge (MA).

Menicucci, Domenico & Pier Luigi Sacco, 1997. Evolutionary dynamics with λ-players. Mimeo, Department of Economics, University of Florence.

Mueller, Dennis. 1986. Rational egoism versus adaptive egoism as fundamental postulate for a descriptive theory of human behavior. Public Choice 51: 3-23.

Ockenfels, Axel & Joachim Weimann, 1999. Types and patterns: an experimental East-West-German comparison of cooperation and solidarity. Journal of Public Economics 71: 275-287.

Palomino, Frédéric & Fernando Vega-Redondo, 1999. Convergence of aspirations and (partial) cooperation in the prisoner's dilemma. International Journal of Game Theory 28: 465-488.

Rapaport, Carol. 1995. Apparent wage discrimination when wages are determined by nondiscriminatory contracts. American Economic Review 85 (5): 1263-1277.

Roth, Alvin. 1995. Bargaining experiments. In A. Roth & J. Kagel (eds.) Handbook of Experimental Economics, Princeton University Press, Princeton.

Rubinstein, Ariel. 1986. Finite Automata Play the Repeated Prisoner's Dilemma. Journal of Economic Theory 39: 83-96.

Sacco, Pier Luigi. 1994. Discussion of Björnerstedt and Weibull's "Nash equilibrium and evolution by imitation". In K.J. Arrow et al. (eds.) Rationality in economics,

Macmillan, London.

Sacco, Pier Luigi & Stefano Zamagni, 1996. An evolutionary dynamic approach to altruism. In F. Farina, F. Hahn & S. Vannucci (eds.) Ethics, rationality and economic behavior, Clarendon Press, Oxford.

Sahlins, Marshall. 1972. Stone Age Economics. De Gruyter, New York.

Schelling, Thomas. 1960. The Strategy of Conflict. Harvard University Press, Cambridge (MA).

Schlag, Karl Hermann. 1994. Why imitate, and if so, how? Mimeo, Department of Economics, University of Bonn.

Schnabl, Wolfgang, Peter Stadler, Christian Forst & Peter Schuster, 1991. Full characterization of a strange attractor. Physica D 48: 65-90.

Sen, Amartya. 1967. Isolation, Assurance and the Social Rate of Discount. Quarterly Journal of Economics 81: 112-125.

Sen, Amartya. 1974. Choice, orderings and morality. In S. Körner (ed.) Practical reason, Blackwell, Oxford.

Smith, Peter & Michael Bond, 1993. Social psychology across cultures. Harvester, Hemel Hampstead.

Smith-Lovin, Lynn. 1993. Can emotionality and rationality be reconciled? Rationality and Society 5: 283-293.

Sugden, Robert. 1991. Rational Choice: A Survey of Contributions from Economics and Philosophy. The Economic Journal 101: 751-785.

Taylor, Michael. 1987. The Possibility of Cooperation. Cambridge University Press, Cambridge (MA).

Taylor, Peter & Leo Jonker, 1978. Evolutionarily Stable Strategies and Game Dynamics. Mathematical Biosciences 61: 51-63.

Weibull, Jörgen. 1995. Evolutionary Game Theory. The MIT Press, Cambridge (MA).

Weimann, Joachim. 1994. Individual behavior in a free riding experiment. Journal of Public Economics 54: 185-200.

Wildavsky, Aaron. 1992. Indispensable framework or just another ideology? Rationality and Society 4: 8-23.

TABLE 1

	NC_{PD}	NC_{AG}	C_{AG}	C_{OR}
NC_{PD}	4	4	8	8
NC_{AG}	5	5	7	7
C_{AG}	3	3	9	9
C_{OR}	2	2	10	10

TABLE 2

	NC_{PD}	NC_{AG}	C_{AG}	C_{OR}
NC_{PD}	1	1	12	12
NC_{AG}	5	5	9	9
C_{AG}	3	3	10	10
C_{OR}	6	6	7	7

TABLE 3

	NC_{PD}	NC_{AG}	C_{AG}	C_{OR}
NC_{PD}	4	4	13	13
NC_{AG}	6	6	7	7
C_{AG}	5	5	12	12
C_{OR}	3	3	13.5	13.5

TABLE 4

	NC_{PD}	NC_{AG}	C_{AG}	C_{OR}
NC_{PD}	1	1	12	12
NC_{AG}	5	4	9	9
C_{AG}	3	$3+\alpha$	10	11
C_{OR}	6	7	8	$8+\beta$